Posted on 12 Mar 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167865036.67121978/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Transient voltage stability assessment based on transfer learning for small samples

Jiayi Zhang[1], Hui Ren[1], Xi Wang[1], zheng zhibin[2], Jinling Lu[1], and Fei Wang[3]

[1]North China Electric Power University - Baoding Campus
[2]State Grid Shandong Electric Power Company
[3]North China Electric Power University

March 12, 2023

Abstract

Large penetration of renewable energy sources into the power grid has increased the complexity of power system operation and greatly reduced the ability of the system to withstand large disturbances. The frequent occurrence of voltage instability problems in the power grid has brought new challenges to the assessment of transient stability analysis of the power system. To achieve fast and accurate assessment of transient voltages, a transfer learning-based transient voltage stability assessment with small samples is proposed, introducing a domain transfer learning approach, embedding a cooperative attention mechanism in the residual network during the feature extraction stage to capture long-range correlations between features, and using adversarial approaches to reduce the differences between samples from different data sets, using the source domain to guide the target domain for network training to improve model's evaluation capability when the number of samples is insufficient, enhance the generalisation performance of the network, and effectively improve the performance of real-time power system transient voltage stability evaluation in the absence of sufficient historical data. Testing on an improved New England 39-node system validates the superiority of this method in transient voltage stability assessment and provides a new approach to practical field transient voltage stability assessment.

Manuscript Details

Article title

Transient voltage stability assessment based on transfer learning for small samples

Running head/short title

Transient voltage stability evaluation by machine learning

Names of all authors in the same order as mentioned in ScholarOne

Jiayi Zhang [1], Hui Ren [1*], Xi Wang [1], Zhibin Zheng [2], Jinling Lu [1], Fei Wang [1]

Author's contribution

Jiayi Zhang: Investigation, Methodology, Software, Validation, Writing – original draft
Hui Ren: Methodology, Conceptualization, Formal analysis, Supervision, Writing – review & editing
Xi Wang: Data curation, Software, Validation, Visualization
Zhibin Zheng: Investigation, Methodology， Resources, Supervision
Jinling Lu: Conceptualization, Resources, Supervision
Fei Wang: Formal analysis, Resources, Supervision

Affiliation of all authors

[1] School of Electrical and Electronic Engineering, North China Electric Power University, 619 Yonghua Street, Baoding, China
[2] State Grid Shandong Electric Power Company Zaozhuang Power Supply Bureau, No.999 Huanghe Road, Zaozhuang, China

Postal and E-mail address of the corresponding author

Postal address: School of Electrical and Electronic Engineering, North China Electric Power University, 619 Yonghua Street, Baoding, China
E-mail address: hren@ncepu.edu.cn (H. Ren)

Funding information (Please mention "Funding: None" if you have not received any support for your research)

Funding: None

Conflict of Interest statement

The author and co-authors of this article have no conflict of interest to disclose.

Permission to reproduce materials from other sources

None

Data Availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Article type

Original Research

Corresponding Author

Hui Ren

Corresponding Author's Institution

North China Electric Power University

Submission Files Included in this PDF

File Name	**[File Type]**
20230307-01-Main Document.pdf	[Main Document]
20230307-02-Highlights.pdf	[Highlights]
20230307-03-Declaration-of-competing-interests.pdf	[Conflict of Interest]
20230307- Cover Letter.pdf	[Cover Letter]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'

Highlights

- A small sample transient voltage stability assessment method is proposed to assess the transient voltage stability of the grid in real time with an insufficient number of samples, enabling enhanced generalization performance of the network.
- The ResNet-Coordinate Attention model is incorporated in the DANN model to enhance the DANN network's ability to capture valid information, and a co-attentive mechanism is embedded in the residual network during the feature extraction stage to capture long-range correlations between features, using an adversarial approach to reduce the variance of samples from different datasets.
- The proposed method is able to use the source domain knowledge as a guide and still has good classification performance in scenarios with few trainable samples, improving the evaluation performance of the network in the presence of insufficient data.

Transient voltage stability assessment based on transfer learning for small samples

Jiayi Zhang [1], Hui Ren [1*], Xi Wang [1], Zhibin Zheng [2], Jinling Lu [1,] Fei Wang [1]

[1] School of Electrical and Electronic Engineering, North China Electric Power University, 619 Yonghua Street, Baoding, China
[2] State Grid Shandong Electric Power Company Zaozhuang Power Supply Bureau, No.999 Huanghe Road, Zaozhuang, China
[*]hren@ncepu.edu.cn (H. Ren)

Abstract: Large penetration of renewable energy sources into the power grid has increased the complexity of power system operation and greatly reduced the ability of the system to withstand large disturbances. The frequent occurrence of voltage instability problems in the power grid has brought new challenges to the assessment of transient stability analysis of the power system. To achieve fast and accurate assessment of transient voltages, a transfer learning-based transient voltage stability assessment with small samples is proposed, introducing a domain transfer learning approach, embedding a cooperative attention mechanism in the residual network during the feature extraction stage to capture long-range correlations between features, and using adversarial approaches to reduce the differences between samples from different data sets, using the source domain to guide the target domain for network training to improve model's evaluation capability when the number of samples is insufficient, enhance the generalisation performance of the network, and effectively improve the performance of real-time power system transient voltage stability evaluation in the absence of sufficient historical data. Testing on an improved New England 39-node system validates the superiority of this method in transient voltage stability assessment and provides a new approach to practical field transient voltage stability assessment.

1. Introduction

With the expansion of the power grid and the operation of hybrid AC and DC transmission lines, the complexity of the operating conditions is increasing, renewable energy sources such as wind power and photovoltaics are extremely volatile and stochastic in form, and the widespread use of renewable energy sources has led to a deterioration of the operating environment of modern power systems and a significant reduction in the system's ability to withstand large disturbances. The safe and stable operation of power systems is becoming increasingly challenging [1,2,3]. Failure to accurately predict the state of the system after a fault may destabilise the system. Therefore, fast and accurate transient stability assessment (TSA) methods for power systems have become an urgent need for modern power systems [4].

In recent years, due to the rapid development of artificial intelligence, deep learning has shown an important role in various fields. As the most advanced technology in the field of language learning, deep learning has the advantages of high abstraction ability, automatic feature extraction and good convergence [5]. Deep learning techniques have been widely used in TSA due to their significant advantages in feature extraction [6]. The literature [7] proposed a data-driven, distributed and easily transferable method for short-term voltage stability assessment, but with slightly lower accuracy and without considering renewable energy sources. The literature [8] proposed a deep neural network model incorporating risk-averse learning, which significantly improved the success rate of emergency load shedding with insignificant increase in load shedding, significantly saved control costs, and improved the efficiency of solving emergency load shedding problems with real-time load shedding decision capability. However, it is only for short-term voltage stability assessment of power systems.

Deep learning models often perform poorly in new tasks, generalise poorly and require new networks to be reconstructed for training. Furthermore, reliable standards for obtaining large-scale, balanced data and accurate labels are difficult and expensive. In addition to data annotation, training data collection itself is costly and laborious in practical applications [7]. Limited by resource constraints and the high cost of manual labelling, the samples available for training models (labelled samples) are not always sufficient, making it difficult for deep learning-based evaluation methods to take advantage of their strengths, exacerbating the difficulty of sample evaluation and accentuating the misclassification problem.

At present, most existing short-term load forecasting methods are studied based on scenarios with sufficient historical load data. When historical load data are insufficient, deep learning for load forecasting is prone to overfitting, and transfer learning is commonly used to deal with the problem of network overfitting when there is a shortage of samples [9], however, the effect of transfer learning is more closely related to the degree of correlation between two data. Therefore, it remains a major challenge for transient voltage stability assessment to reduce the impact of sample distribution differences on assessment performance and to improve the generalisation ability of the model.

Literature [9] proposed a short-term power load prediction model based on mixing and transfer learning. In the case of insufficient historical load data, the data set is extended by combining mixing and transfer learning to solve the problem that new houses lack historical load data to establish a reliable load prediction model. However, it is only applicable to the load prediction of a single house. It is not suitable for large power system. Literature [10] proposed a deep neural network transfer learning method based on depth representation to deal with the migration scenario of passive domain data. This method adjusts all parameters of the pre-training network by minimizing the empirical error of the target domain data, fully learns the specific representation of the target domain, transfers the domain knowledge from the source depth neural network to the target domain, and improves the generalization ability of the target network. However, it only applies to single-layer transport embedding. In literature [11], data enhancement is adopted to increase the size of training data set, and transfer learning is adopted to reduce the amount of computation involved in training. However, the transfer-learned VGG-19 does not perform very well at shallow layers. Literature [12] scattering convolution network was proposed based on Time-Scattering Convolutional Network(TScatNet) cross-domain diagnosis model, which is based on the nonlinear Morlet wavelet convolution kernels and average pooling layer to extract domain invariant features, The error caused by domain transfer is eliminated, and the network hyperparameter adjustment is simple and explanatory.

To solve the above problems, in order to improve the evaluation ability of the model when the number of samples is insufficient and enhance the generalization performance of the network, this paper proposes an intelligent method of transient voltage stability evaluation based on domain transfer learning, which can evaluate the stability state of the power grid in real time. This method integrated two models, DANN and ResNet-Coordinate Attention. In the stage of feature extraction, coordinated attention mechanism was embedded in the residual network to capture the long-distance correlation between features, and countermeasures were adopted to narrow the differences between samples of different data sets. The domain transfer method based on DANN obtains the characteristics of voltage operation in the transient process from the scene with complete samples (source domain) and guides the scene with small sample data (target domain), so as to improve the evaluation performance of the network in the case of insufficient data. Finally, the improved New England 39-node system is used to verify the applicability of the proposed method.

The arrangement of this paper is as given in Fig. 1. The proposed transient voltage early warning framework consists of two parts: off-line training stage and online application. The off-line training part is composed of data set generation module and transient evaluation model classifier. In the generation stage of the data set, the data set of transient voltage stability can be generated through historical data or simulation of the system under different conditions of large disturbance, and the data can be pre-processed through normalization to reduce the impact of differences between features on the model evaluation performance. In the model training stage, the evaluation model learns the implicit relationship between the system state and the transient voltage dynamic process through the dynamic information of the samples obtained from the data set. In the online evaluation stage, the real-time measurement data of the system is input into the trained evaluation model to evaluate the transient voltage stability of the system in real time.

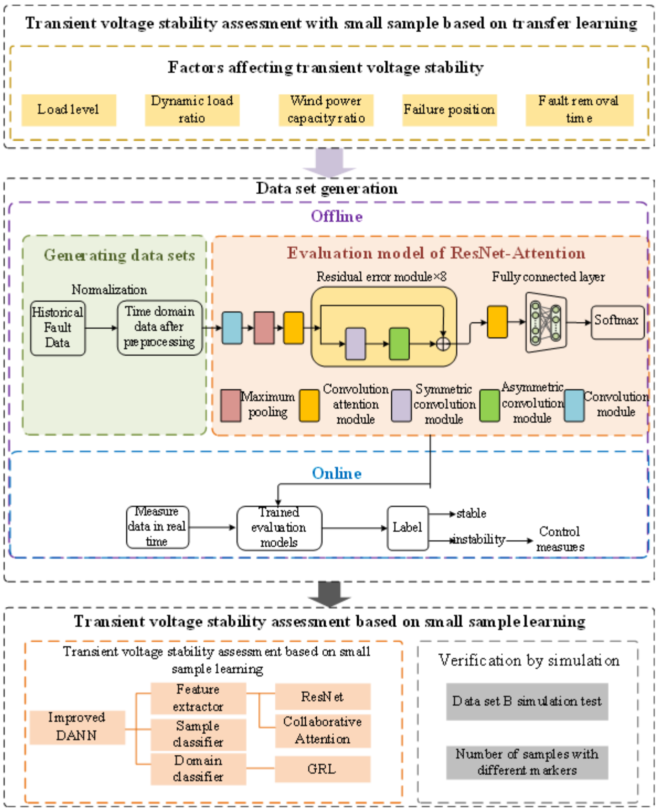


Fig. 1. *Work arrangement of this paper*

2. Factors affecting transient voltage stability of the system

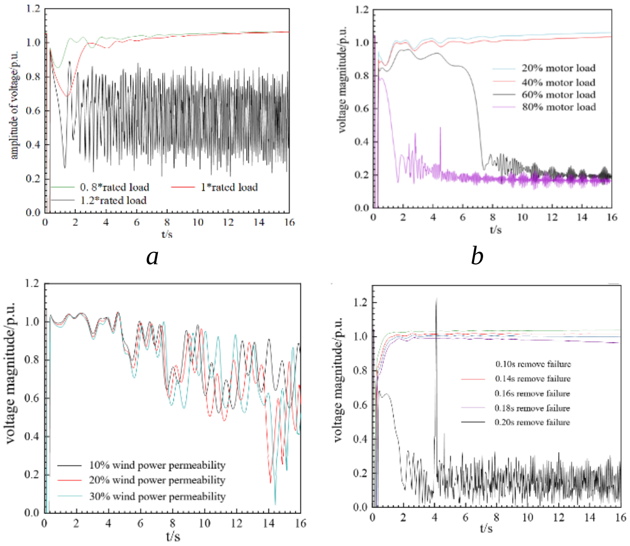
Transient voltage stability refers to the ability of the system to be disturbed without voltage collapse. The duration of transient voltage instability is usually less than 10s. Because transient voltage instability is not easily detected in the early stage, it is difficult for operators to accurately judge the system state and quickly carry out emergency control in the early stage of fault occurrence, which leads to further deterioration of the accident and brings severe challenges to the safety and stability assessment of power grid. Therefore, in real-time monitoring, if the transient voltage instability phenomenon can be accurately predicted and effective methods can be adopted quickly, it is of great significance to curb voltage collapse and chain reaction after large disturbance, and improve the safety and stability of power grid.

Through conventional analysis, system may undergo transient voltage instability when the following factors at the bus observed vary: 1) load level; 2) dynamic load ratio; 3) the penetration of wind power of the system; 4) the position of the failure, and 5) the clearing time of the failure.

We take IEEE-39 bus as the test system, and by changing the above mentioned five factors, 4025 test cases are designed and 3960 samples are collected. Simulation results are given in Fig.2 and Fig.3.

- Base case: rated load, the removal time to be 0.2s, no wind power generation is injected. no dynamic load is considered. Failure position is chosen to be at the end of the line.
- Group 1: the load level is chosen to be 0.8*rated load, 1*rated load, 1.2*rated load.
- Group 2: Motor is chosen as an example of dynamic load [13], and the ratio of dynamic load at each bus is chosen to be 20%, 40%, 60% and 80%, respectively.

- Group 3: Wind power penetration ratio is chosen to be 10%, 20% and 30%, respectively.
- Group 4: Failure removal time to be 0.1s, 0.14s, 0.16s, 0.18s and 0.20s, respectively.



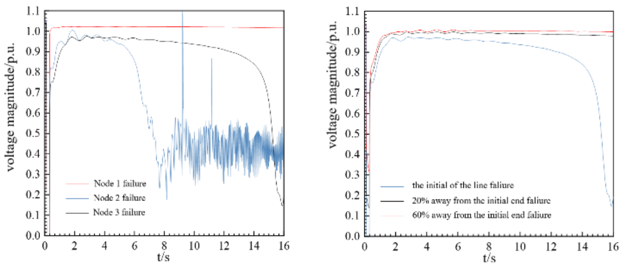
a *b* *c* *d*

Fig. 2. *Voltage curves of IEEE-39 system under different conditions*

(a) Voltage curve under different load levels, **(b)** Voltage curve for different motor ratios, **(c)** Voltage curve under different wind power penetration ratio, **(d)** Voltage curve under different fault removal time

In order to analyse the voltage characteristics of the system at different fault points, the voltage curves of node 3(the head of the branch) at different nodes and at different locations of the branch between node 3 and node 4 are taken as examples to briefly describe the voltage curves at different fault locations。

- Group 1: Failure position is designed to be node 1, 2 , 3, respectively.
- Group 2: The failure position of a line is chosen to be at the initial of the line, 20% away from the initial end, and 60% from the initial end, respectively.



a *b*

Fig. 3. *Voltage curve of IEEE-39 system under different node faults*

(a) Voltage curve under different node faults, **(b)** Voltage curves of branches at different positions between node 3 and node 4 faults

This section takes the IEEE39-node system as an example to simulate the above five different fault scenarios and quantitatively analyse the factors affecting the transient voltage instability of the power grid, which provides a basis for the subsequent diversification selection of transient voltage stability evaluation network samples.

3. Classification Algorithm Introduction

3.1. Transient voltage stability evaluation process based on small sample learning

In the actual system, the training data acquisition itself has the problems of high cost and heavy workload in the practical application, limited by resources and the high cost of manual labeling, so the samples available for training models have the problem of insufficient sample quantity.In the case of insufficient samples, the capability of deep-learning algorithm to capture features needs to be enhanced to fully explore the implicit correlation between dynamic curve and transient voltage stability in the transient process.

In this paper, domain transfer is used to solve the problem. Due to reduce the impact of the difference between source domain and target domain, and to deal with the shortcomings of domain adaptation in transfer learning, Domain-Adversarial Training of Neural Networks(DANN) is proposed[14], In this paper, the Deep-learning model DANN is adopted to realize domain transfer, and the deep evaluation model ResNet-Coordinate Attention is trained on a large number of sample sets after migration to improve the performance of the model with small samples.

The algorithm is mainly divided into three stages: domain transfer, model training and online assessment.

(1) In the domain transfer stage, the deep-learning evaluation model based on ResNet-Coordinate Attention was used to extract the features of the source domain and target domain, train the parameters of the classifier with the samples of the source domain, and reduce the difference of the evaluation of the source domain and target domain by the network in an adversarial way.

(2) In the off-line model training stage, the data in the source domain during the process of domain transfer is used as the sample support to learn the knowledge of the source domain and improve the performance of the model in the target domain.

(3) In the online evaluation stage, the evaluation model with the best performance in the training process is used in the actual system to evaluate the real-time state of the system and judge the transient voltage stability state of the system.

The transient voltage warning framework proposed in this paper consists of two parts: off-line training stage and online application. The off-line training part is composed of data set generation module and transient evaluation model classifier. In the generation stage of the data set, the data set of transient voltage stability can be generated through historical data or simulation of the system under different conditions of large disturbance, and the data can be pre-processed through normalization to reduce the impact of differences between features on the model evaluation performance. In the model training stage, the evaluation model learns the implicit relationship between the system state and transient voltage dynamic process through a large amount of dynamic information in the data set. In the online evaluation stage, the real-time measurement data of the system is input into the trained evaluation model to evaluate the transient voltage stability of the system in real time.

3.2. Small sample stability assessment based on transfer learning

3.2.1 Attention Mechanism: Attention Mechanism (AM) is derived from the visual information capture mechanism. When acquiring information, AM will automatically capture the region of attention. In deep-learning scenarios, AM is commonly used to calculate the critical degree of each feature and give the key information a large weight to improve the efficiency and accuracy of the network.

The early use of AM in the field of semantic segmentation can be divided into two stages: weight generation and attention formation. In the weight generation stage, the weight O_l of the *lth* feature I_l can be expressed as follows.

$$O_l = AM(I_l) = soft\,max(\,h(I_l, C_l)) \tag{1}$$

Where, AM is the attention weight generation operation, C_l is the value of other elements corresponding to I_l. $h(\cdot,\cdot)$ is the correlation operation, the common way is a little product operation, addition operation and so on. In the Attention Formation Stage, the weighted average method is used to summarize the input information:

$$Att(I) = \sum_{l=1}^{M}(O_l \times U_l) \tag{2}$$

In Equation (2), $Att(I)$ is the attention of input vector I, O_l is the weight of feature I_l in the weight generation stage, I_l is the value of U_l, and M is the number of elements in vector I.

In classification problems, AM is used to capture the key elements in features that affect classification decisions, including spatial attention [15] and channel attention [16]. Nowadays, many scholars have explored the networks based on AM architecture, giving birth to networks such as Transformer [17] and BERT [18].

3.2.2 Collaborative attention mechanism: In channel-based attention such as CBAM and SENet(Squeeze and Excitation Networks), global pooling is commonly used to encode global spatial information. However, the above attention can only capture local information, and it is difficult to mine feature dependencies in a long range. In order to further enhance the ability of the network to extract the dependency relationship of long-distance data, the Coordinate Attention mechanism (CA) [20] is introduced. The attention module adds the location information to CAM. The decomposition channels are two one-dimensional encoders along different directions to efficiently integrate the spatial coordinate information into the feature map, so as to give different weights to each channel according to the importance of features and enhance the sensitivity of the model to the information. In Resnet-Coordinate Attention structure, Coordinate Attention is located at the first feature extraction stage and the final output stage respectively. The CA structure is shown in Fig. 4:

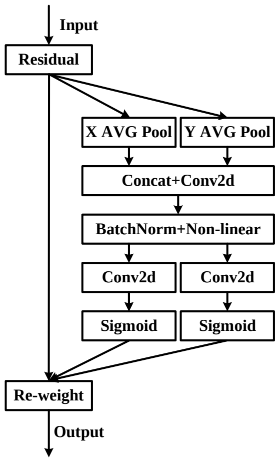


Fig. 4. *Collaborative attention structure*

Fig. 4 shows the structure diagram of collaborative attention module. CA module encodes channel correlation and long-range correlation with accurate position information, which can be specifically divided into two stages: coordinate embedding and attention generation.

(1) Coordinate embedding stage

The collaborative attention module replaces the global pooling with two pooling kernels of size (*L*,*1*) and (*1*,*M*) to carry out the feature encoding process, so that the network can capture the spatial long-range dependence relationship with accurate location information. For the input feature information, a pooling kernel of length *L* and a pooling kernel of width *M* are used to pool each channel in the horizontal and vertical directions respectively. So far, the output result of the *kth* channel can be expressed as follows:

$$Z_k = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} x_k(i,j) \tag{3}$$

Where, Z_k is the input feature, x_k is the output result after the information embedding stage, *L* and *M* are the length and width of two pooling kernel channels, respectively. The coding process of features in horizontal and vertical directions can be expressed by Equations (4) and (5) :

$$Z_c^l(l) = \frac{1}{M}\sum_{0 \le i < M} x_c(l,i) \tag{4}$$

$$Z_c^m(m) = \frac{1}{L}\sum_{0 \le j < L} x_c(j,m) \tag{5}$$

The two encoding processes fuse different features in perpendicular directions to generate a set of abstract features with location information. The two transformations enable the attention module to obtain the long-range correlation between information in one direction and preserve the precise location information obtained in the other direction, which is convenient for the network to capture the key nodes in the transient process.

(2) Attention generation stage

The collaborative attention generation stage aims to make full use of the location information, accurately locate the key feature information, and effectively capture the deep correlation between channels. In the attention generation stage, the two transformed output results are firstly feature concatenated, and then the convolution operation is carried out with a convolution kernel of size *1*, and its dimension is reduced. Secondly, the features after dimensionality reduction are batch normalized to get F_1, and then the Sigmoid function is used to generate the feature *f*, namely:

$$f = \sigma(F_1([z^l, z^m])) \tag{6}$$

Where, $[\cdot,\cdot]$ is the process of stitching coding results, σ is the sigmoid activation function, and f is the mapping of spatial features after coding in horizontal and vertical directions respectively.

For feature f, it is decomposed into f^l land f^m , and then, the feature f^l land f^m are convolution transformed to get the feature map with the same size as the original input g^h、 g^w,

$$g^l = \sigma(K_l(f^l)) \tag{7}$$

$$g^m = \sigma(K_m(f^m)) \tag{8}$$

Where, σ is sigmoid activation function, g^l and g^m are the attention weights of features in vertical and horizontal directions, respectively. And K_h are K_w the 1×1 convolution operations for the features f^l and f^m, respectively. The output of CA module can be expressed as formula (9):

$$y_c(i,j) = x_c(i,j) \times g_c^l(i) \times g_c^m(j) \tag{9}$$

3.2.3 Improve DANN network:

DANN will generate Generative Adversarial Nets(GANs)[20] the sample generator replace convolution kernels to extract features, without altering discrimination on the basis of and behind the feature extractor to join FCL to accomplish the classification task. The structure of DANN is shown in Fig. 5.

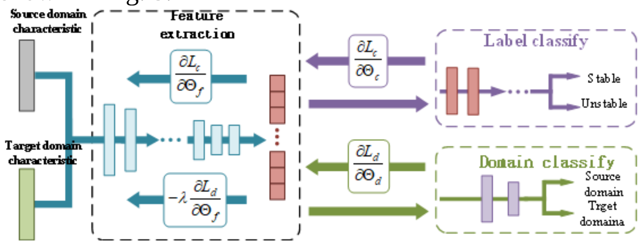


Fig. 5. *Structure diagram of improved DANN*

As shown in Fig. 5, DANN consists of three components: Feature Extractor (FE)G_f, Label Predictor (LP)G_c and Domain Classifier (DC)G_d.

The feature extractor uses ResNet-Coordinate Attention model, its structure is similar to ResNet-Attention network structure, which replace the original CBAM module with Coordinate Attention. In the feature extraction stage, the network captures the deep feature F_h in the input information x_k through the feature extractor:

$$F_h = G_f(x_k; \Theta_f) \tag{10}$$

After feature extraction, data need to be classified by label classifier G_c , and the classification loss of samples L_c can be expressed as:

$$L_c = \sum_{k=1}^{n} L_c(G_c(G_f(x_k)), y_k) \tag{11}$$

In Equation (11), x_k is the *kth* sample, y_k is the actual label of sample x_k, and n is the number of samples. In order to reduce the difference between the distribution of sample x_s in the source domain and sample x_t in the target domain, DANN performs domain regularization on the output result $G_f(f(x))$ obtained in the feature extraction stage. The domain regularization process uses the domain discriminator G_d to identify the domain belonging of samples.

$$F_d = G_d(G_f(x);\ \Theta_d) \tag{12}$$

In Equation (12), F_d is the judgment result of the domain discriminator, and its value is 0 or 1, representing the source domain and target domain respectively. Θ_d is the parameter of the domain classifier. In this case, the loss function introduced by the domain discriminator is:

$$L_d = \sum_{k=1}^{n} L_d(G_d(G_f(x_k)), y_k) \tag{13}$$

In summary, the loss function of DANN is defined as：

$$E(\Theta_f, \Theta_c, \Theta_d) = \sum_{k=1}^{N} L_c(G_c(G_f(x_k; \Theta_f); \Theta_c), y_k)$$

$$-\lambda \sum_{k=1}^{N} L_d(G_d(G_f(x_k; \Theta_f); \Theta_d), y_k)$$

$$= \sum_{k=1}^{N} L_c^k(\Theta_f, \Theta_c) - \lambda \sum_{k=1}^{N} L_d^k(\Theta_f, \Theta_d) \tag{14}$$

$E(\Theta_f, \Theta_c, \Theta_d)$ is the loss function of the DANN network, L_c is the loss of label classification, L_d is the domain classification loss, and the λ parameter is the tradeoff coefficient of the model between the two objectives. When the DANN parameter reaches the optimum, the following conditions should be met: 1) In order to enhance the classification accuracy of the model classifier, it is necessary to change the domain classifier parameter to maximize the ; 2) In order to solve the problem of domain adaptability, it is necessary to adjust the feature extractor and label classifier parameters Θ_f and Θ_c to minimize L_c :

$$(\hat{\Theta}_f, \hat{\Theta}_c) = arg \min_{\Theta_f, \Theta_c} E(\Theta_f, \Theta_c, \hat{\Theta}_d) \tag{15}$$

$$\hat{\Theta}_d = arg \max_{\Theta_d} E(\hat{\Theta}_f, \hat{\Theta}_c, \Theta_d) \tag{16}$$

$\hat{\Theta}_f$,$\hat{\Theta}_c$ and $\hat{\Theta}_d$ are the values of G_f,G_c and G_d, when the DANN loss function E is minimized, respectively. In order to avoid the same as GAN structure using lock generator and discriminant method of parameters in stages training, DANN joined the Gradient Reversal Layer (GRL) in the network structure. In forward propagation, GRL plays the role of information transmission and does not change any parameters. While during backpropagation, GRL obtains the gradient from the back network, multiplies the gradient with $-\lambda$ and passes it forward:

$$\Theta_f \leftarrow \Theta_f - \mu\left(\frac{\partial L_c^k}{\partial \Theta_f} - \lambda \frac{\partial L_d^k}{\partial \Theta_f}\right) \tag{17}$$

$$\Theta_c \leftarrow \Theta_c - \mu \frac{\partial L_c^k}{\partial \Theta_c} \tag{18}$$

$$\Theta_d \leftarrow \Theta_d - \mu \frac{\partial L_d^k}{\partial \Theta_d} \tag{19}$$

In Equations (17) to (19), μ is the learning rate, and the gradient inversion layer is defined as the pseudo-function $R_\lambda(x)$, then:

$$R_\lambda(x) = x \tag{20}$$

$$\frac{dR_\lambda}{dx} = -\lambda I \tag{21}$$

In Equations (20) to (21), I is the identity matrix, then the loss function can also be expressed as:

$$E(\Theta_f, \Theta_c, \Theta_d) = \sum_{k=1}^{N} L_c(G_c(G_f(x_k; \Theta_f); \Theta_c), y_k)$$

$$+ \sum_{k=1}^{N} L_d(G_d(R_\lambda(G_f(x_k; \Theta_f)); \Theta_d), y_k) \tag{22}$$

3.3. Evaluation Index

Transient voltage stability evaluation is a dichotomous problem. The accuracy rate, false alarm rate, false alarm rate and F1-score are used to comprehensively evaluate the deep-learning network. The confusion matrix of transient voltage stability evaluation is shown in Table 1.

Table 1 Confusion matrix for transient voltage stability evaluation

Sample Label	Evaluation Label	
	Stable	Unstable
Stable	n_{TP}	n_{FN}
Unstable	n_{FP}	n_{TN}

In Table 1,n_{TP} represents the samples whose actual tag is stable and the predicted result is stable; n_{FN} represents the samples whose actual tag is stable and the predicted result is unstable; n_{FP} represents the samples whose actual tag is voltage instability and the predicted label is stable; n_{TN} represents the samples whose actual tag is transient voltage instability and the predicted sample is voltage instability.

3.3.1 Accuracy: Accuracy can clearly and intuitively judge the performance of the model, which is often used as an index to evaluate classification problems. The formula of Accuracy(ACC) is as follows:

$$Accuracy = \frac{n_{TP}+n_{TN}}{n_{TP}+n_{FP}+n_{FN}+n_{TN}} \tag{23}$$

3.3.2 False alarm rate and false alarm rate: The false alarm rate F_D represents the model's ability to correctly evaluate the transient voltage instability samples. The smaller the alarm rate F_D is, the stronger the model's ability to evaluate the transient voltage instability samples is. The false alarm rate F_A represents the evaluation ability of the model to accurately predict stable samples, and the smaller the false alarm rate is, the stronger the evaluation ability of the model for unstable samples is. The false alarm rate and false alarm rate are shown in Equations (24) and (25)

$$F_D = \frac{n_{FP}}{n_{FP}+n_{TN}} \tag{24}$$

$$F_A = \frac{n_{FN}}{n_{FN}+n_{TP}} \tag{25}$$

3.3.3 F1-score: F1-score is an index used to measure the accuracy of classification models in statistics, which can be used to comprehensively evaluate the evaluation ability of models for stable samples and unstable samples. F1-score ranges from 0 to 1. The larger the value, the better the network classification performance. The formula of F1-score is as follows:

$$F1-score = \frac{2\times\frac{n_{TP}}{n_{TP}+n_{FP}}\times\frac{n_{TP}}{n_{TP}+n_{FN}}}{\frac{n_{TP}}{n_{TP}+n_{FP}}+\frac{n_{TP}}{n_{TP}+n_{FN}}} \tag{26}$$

4. Example analysis

In this paper, the improved New England 39-node system is adopted, and node 1 is selected as the balanced node. The number of generators and nodes in the system is 10 and 39, among which there are 19 load nodes, 10 transformer branches and 34 transmission lines. The wind turbine is connected to 35 nodes through transformers. The sixth-order model is selected for the generator, and the wind power permeability is 5%. The simulation model is shown in Fig. 6.

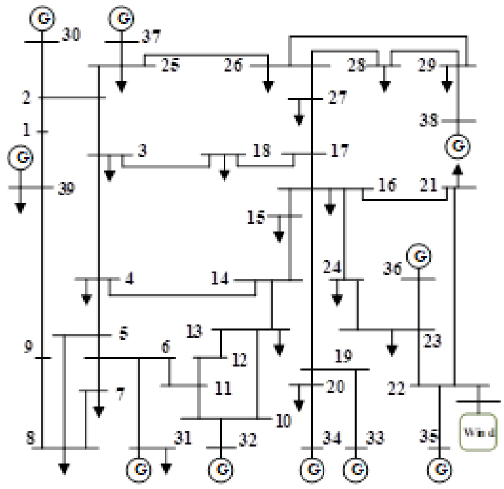


Fig. 6. *Improving the 39-node system in New England*

Through different simulations, datasets A and B are generated. In the failure scenario of data set A, the system load level is set as follows: The system load level ranges from 70% to 115% and is evenly divided into 10 load levels. The load on each bus is composed of constant power load and induction motor load, and the proportion of induction motor load in simulation is 30% and 50%. The time of fault occurrence is set at 0.1s, and the time of fault resection is set at 0.2s, 0.24s, 0.26s, 0.28s, and 0.3s, respectively. The system frequency is 50Hz, and the sampling period is 0.02s. Excluding the samples of transient power Angle instability and island formation, the total number of samples is 5029, the number of stable samples is 2930, and the number of unstable samples is 2099.Label the samples with voltage instability as 1, and set the stable samples as 0.Considering the vigorous development of new energy and system development planning and other factors, set data set B for comparison and verification: By replacing some generators of nodes 30, 32, 33, 34, 35, 36 and 37 with wind turbines, the overall penetration rate of new energy reached 20%. The selection of fault points, fault duration and other conditions were the same as those of dataset A. A total of 4156 samples were generated, among which the number of stable samples: the number of unstable samples was 3507:649.

4.1. Model performance evaluation

In order to evaluate the performance of each model, the simulation is carried out on dataset B, and the indicators of each model are as follows:

Table 2 Simulation results of each model in dataset B

Model	ACC/%	F_D	F_A	F1-score
SVM	82.93	1.0000	**0.0000**	-
CNN	90.02	0.5000	0.0174	0.6311
Light GBM	93.39	0.2746	0.0232	0.7893
RF	93.03	0.3380	0.0145	0.7642
ResNet	94.30	0.2234	0.0262	0.8096
Proposed method	**98.08**	**0.0704**	0.0087	**0.9429**

As can be seen from Table 2, thanks to the DANN module, this algorithm reduces the impact of reducing the model evaluation performance due to the distribution differences of data sets in the adversarial process, and can show excellent generalization performance in the case of large data differences.

Fig. 7 shows the distribution of data sets A and B and their classification effects on the deep-learning model.

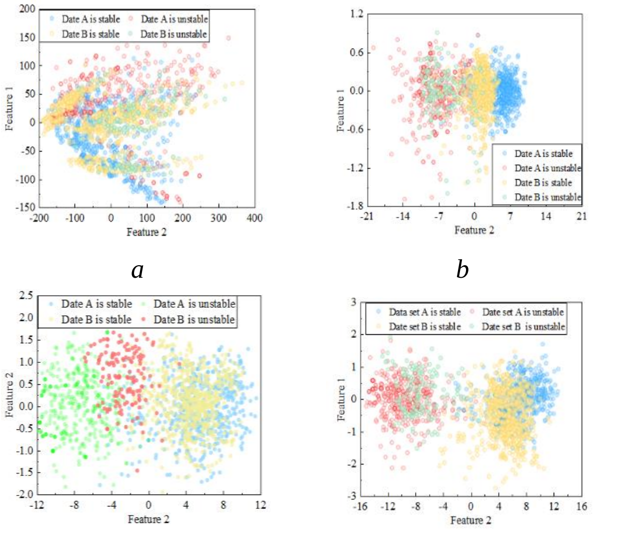


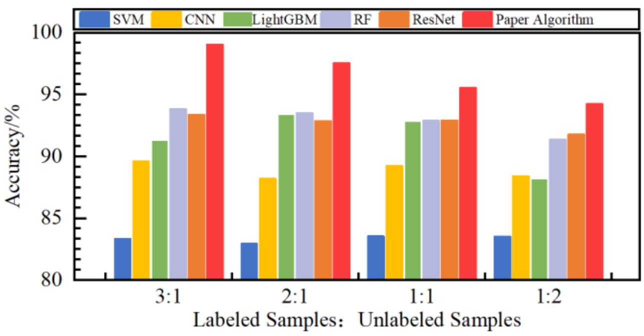
Fig. 7. *Dimensionality reduction results and classification effect of dataset A and B*

(a) Dimensionality reduction of original data, **(b)** ResNet The output of the last layer of the network, **(c)** Improve the output of the last layer of ResNet network, **(d)** Improve the output of the last layer of DANN network

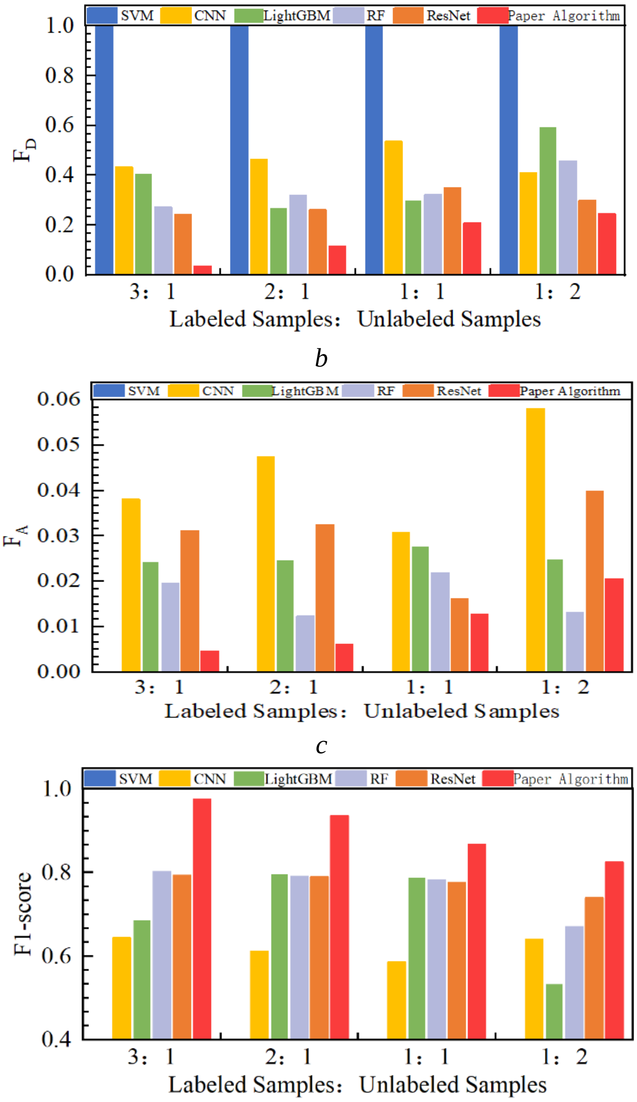
It can be seen from Fig. 7(a) that there are obvious differences in the distribution rules of unstable samples between datasets A and B, which aggravate the misclassification of unstable samples by each model (Fig. 7(b) and (c)).As can be seen from Fig. 7(b) and (c), after the training of the model based on deep-learning, the characteristics of unstable samples in dataset A and stable samples in dataset B are relatively close, which is an important reason for the poor performance of the model on the new dataset. Fig.7(d) shows that the model can effectively reduce the differences between samples in different fields and effectively improve the generalization of the model

4.2. Influence of the number of labelled samples on model performance

In order to further verify the performance of the proposed network under small samples, this section simulates the evaluation performance of each model with different number of labelled samples in dataset B, and the evaluation indexes of each model are shown in Fig. 8.



a



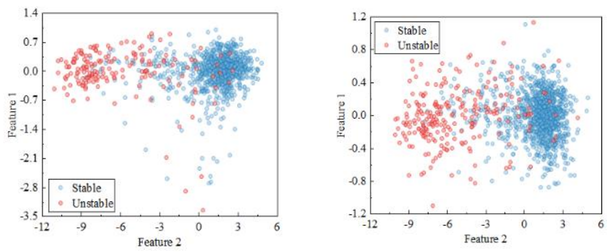
d

Fig. 8. *Indicators of each model with different number of labelled samples*

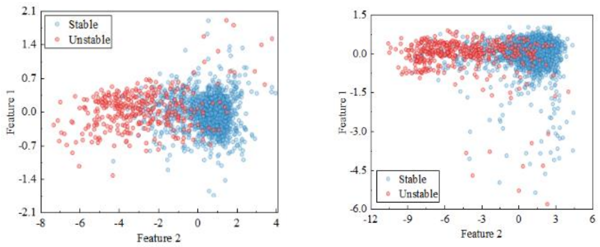
(a) Accuracy, **(b)** F_D, **(c)** F_A, **(d)** F1-score

As shown in Fig. 8, as the number of labelled samples decreases, the evaluation performance of each model decreases. The algorithm in this paper can show excellent evaluation performance when the number of samples is insufficient due to the introduction of source domain knowledge.

Fig. 9 to 11 directly show the classification performance of each model with different number of labelled samples.

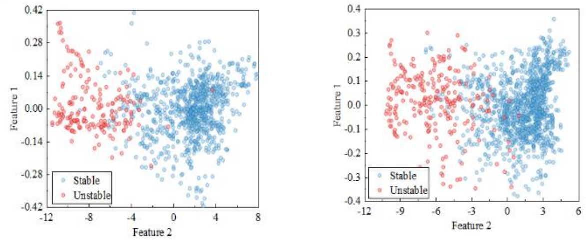


a *b*

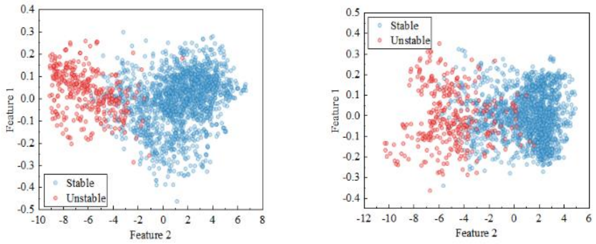


c *d*

Fig. 9. *Classification results of ResNet network under different labelled sample numbers*
(a) 3:1, **(b)** 2:1, **(c)** 1:1, **(d)** 1:2

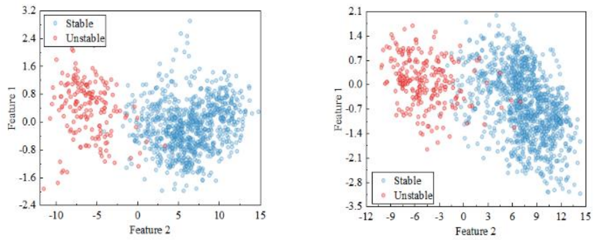


a *b*

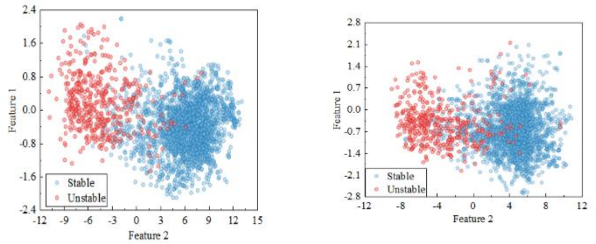


c *d*

Fig. 10. *Classification results of improved ResNet network with different number of labelled samples*
(a) 3:1, **(b)** 2:1, **(c)** 1:1, **(d)** 1:2



a *b*



c *d*

Fig. 11. *Classification results of the improved DANN network with different number of labelled samples*
(a) 3:1, **(b)** 2:1, **(c)** 1:1, **(d)** 1:2

In Fig. 9-11, 3:1 and 1:1 represent the ratio of the number of labelled samples to the number of unlabelled samples. It can be seen that when the number of labelled samples decreases, the distribution of the two types of samples gradually converges in the network classification layer, which is due to the lack of available samples in the training stage, and the model cannot accurately fit the classification boundary of the two types of samples. The algorithm in this paper effectively makes use of the samples in dataset A, and the classification effect can be maintained at A high level even when the number of available samples is insufficient.

5. Summary of this chapter

In this paper, a small sample transient voltage stability evaluation based on transfer learning is proposed. The collaborative attention mechanism is used to accurately locate the key feature information and effectively capture the deep correlation between channels. The sample generator in the DANN generative adversarial network is replaced by a convolution kernel for feature extraction. At the same time, FCL is added to the feature extractor to complete the classification task. Finally, the simulation is carried out on the improved New England 39-node standard system, and the following conclusions are obtained:

1) In the feature extraction stage, residual neural network with collaborative attention mechanism is adopted to capture the dependency relationship of long-distance features and enhance the feature extraction ability of the model. In the training stage, the difference between the data distribution of the source domain and the target domain is narrowed by adversarial method, and the knowledge of the source domain is used to guide the evaluation task of the model in the new scene.

2) The traditional convolutional network is replaced with ResNet model in order to strengthen the ability of DANN network to capture effective information. In this paper, the collaborative attention module is embedded in the network, which can capture the long-scale dependency relationship between features. The improved IEEE39 node system with one-fifth of the wind energy is tested, and the simulation results of each model are compared under different number of labelled samples, which further verifies the effectiveness of the proposed algorithm in the case of insufficient sample data.

3) Aiming at the problem of network overfitting and domain adaptability in transfer learning under small samples, the method proposed in this paper can be guided by the source domain knowledge and still have good classification performance in the scenario with few trainable samples.

6. References

[1] W. Wenjing, X. Jianing, S. Xin, L. Zhihui, et al., 'The research on correlation of electrical power system stability in high proportion wind power area', 2021 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS), Shenyang, China, 2021, pp. 563-567

[2] He P, Wen B, 'Wang H. Decentralized Adaptive Under Frequency Load Shedding Scheme Based on Load Information', IEEE Access 2019;7:52007–14.

[3] G. YANG, J. WANG, D. WANG, S. ZHANG, Y. LI and Y. XU, 'Study on Operation Characteristics of Wind Power Grid-connected Doubly-Fed AC/DC Hybrid System', 2019 IEEE 8th International Conference on Advanced Power System Automation and Protection (APAP), Xi'an, China, 2019, pp. 1804-1808

[4] Javan DS, Rajabi Mashhadi H. 'Wide-area security assessment based on informative variables of power system', Int J Electr Power Energy Syst 2021,129, pp. 0142–615.

[5] Li, X. , et al. 'Deep learning-based transient stability assessment framework for large-scale modern power system', International Journal of Electrical Power & Energy Systems 2022, 139,108010-..

[6] Wang H , Wu S . 'Transient stability assessment with time-adaptive method based on spatial distribution', Electrical Power and Energy Systems, 2022, 143:10846.

[7] Li Y , Zhang M, Chen C . 'A Deep-Learning intelligent system incorporating data augmentation for Short-Term voltage stability assessment of power systems', Applied Energy, 2022, 308:118347.

[8] Liu J , Zhang Y , Meng K . 'Real-time emergency load shedding for power system transient stability control: A risk-averse deep learning method[J]. Applied Energy, 2022, 307:118221.

[9] Lu Y , Wang G , Huang S . 'A short-term load forecasting model based on mixup and transfer learning', Electric Power Systems Research, 2022, 207:107837-.

[10] Tao Y, Xia Y , Ning M . 'Deep representation-based transfer learning for deep neural networks', Knowledge-Based Systems, 2022, 253,109526.

[11] Gao Z , Zhang Y , Li Y . 'Extracting Features from Infrared Images using Convolutional Neural Networks and Transfer Learning', Infrared Physics & Technology, 2020, **105**,(6),103237.

[12] Liu C, Qin C, Shi X, et al. 'TScatNet: An interpretable cross-domain intelligent diagnosis model with antinoise and few-shot learning capability[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 70: 1-10.

[13] A. K. Barnes, S. Basu and A. Mate, 'Dynamic State Estimation-Based Protection for Induction Motor Loads', 2022 North American Power Symposium (NAPS), Salt Lake City, UT, USA, 2022, pp. 1-5

[14] Y. Li et al., 'Channel Discrepancies Adaptive Modulation Recognition Using Domain Adversarial Training', 2021 Asia-Pacific International Symposium on Electromagnetic Compatibility (APEMC), Nusa Dua - Bali, Indonesia, 2021, pp. 1-4

[15] Zhu X, Cheng D, Zhang Z, et al. 'An empirical study of spatial attention mechanisms in deep networks[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 6688-6697.

[16] Hu J, Shen L, Sun G. 'Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[17] Vaswani A, Shazeer N, Parmar N, et al. 'Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[18] Devlin J, Chang M W, Lee K, et al. Bert， 'Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.

[19] Hou Q, Zhou D, Feng J. 'Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 13713-13722.

[20] Goodfellow I, Pouget-Abadie J, Mirza M, et al. 'Generative adversarial nets[J]. Advances in neural information processing systems, 2014, 27.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/ personal relationships which may be considered as potential competing interest: